Supplementary Information (SI) for Journal of Materials Chemistry A.
This journal is © The Royal Society of Chemistry 2025

Supporting Information

Machine Learning-driven Gait-assisted Self-powered Wearable Sensing: A Triboelectric Nanogenerator-based Advanced Healthcare Monitoring

Parag Parashar[a], Manish Kumar Sharma[a,b], Bishal Kumar Nahak[a], Arshad Khan[a,c], Wei-Zan Hsu[d], Yao-Hsuan Tseng[d], Jaba Roy Chowdhury[a], Yu-Hui Huang[a], Jen-Chung Liao[e,*], Fu-Cheng Kao[e,f,*], and Zong-Hong Lin[a,*]

[a]Department of Biomedical Engineering, National Taiwan University, Taipei 10617, Taiwan.
[b]Department of Materials Science and Engineering, National Tsing Hua University, Hsinchu 30013, Taiwan.
[c]International Intercollegiate PhD Program, National Tsing Hua University, Hsinchu 30013, Taiwan.
[d]Institute of Biomedical Engineering, National Tsing Hua University, Hsinchu 30013, Taiwan
[e]Department of Orthopaedic Surgery, Spine Section, Chang Gung Memorial Hospital, Taoyuan 333, Taiwan
[f]College of Medicine, Chang Gung University, Taoyuan 333, Taiwan

[*]Corresponding author email: jcl1265@adm.cgmh.org.tw, b9102034@cgmh.org.tw, zhlin@ntu.edu.tw

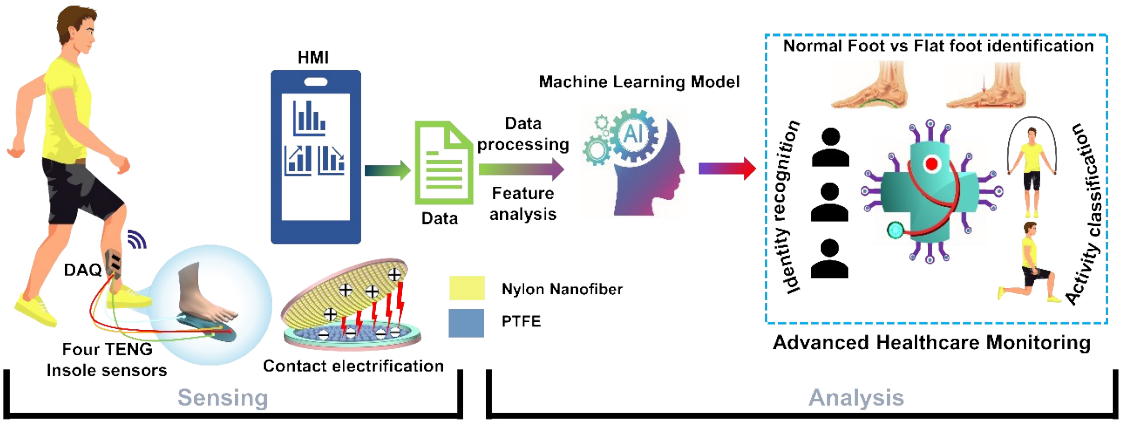


Fig. S1. ML-driven gait-assisted advanced health monitoring system: From sensing to analysis.

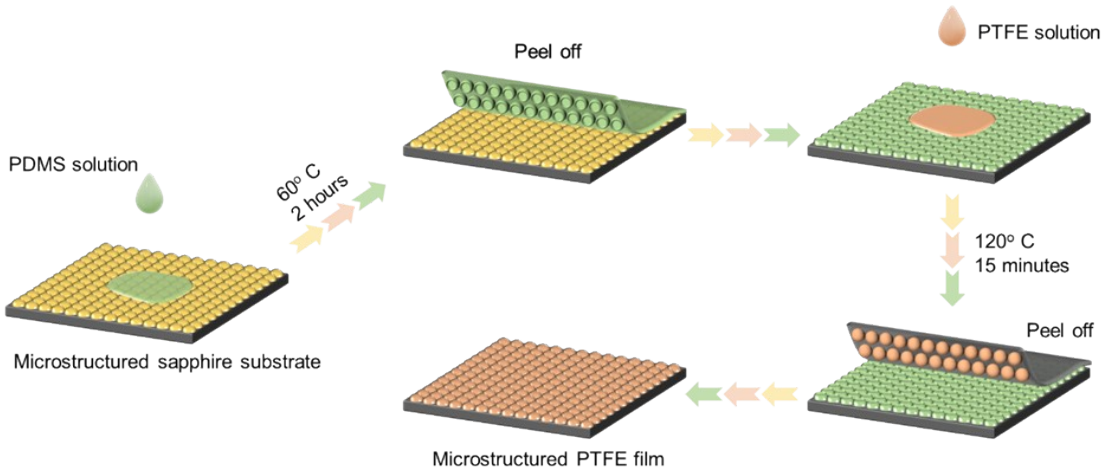


Fig. S2. Schematic of double replica molding for replicating the microstructured PTFE film.

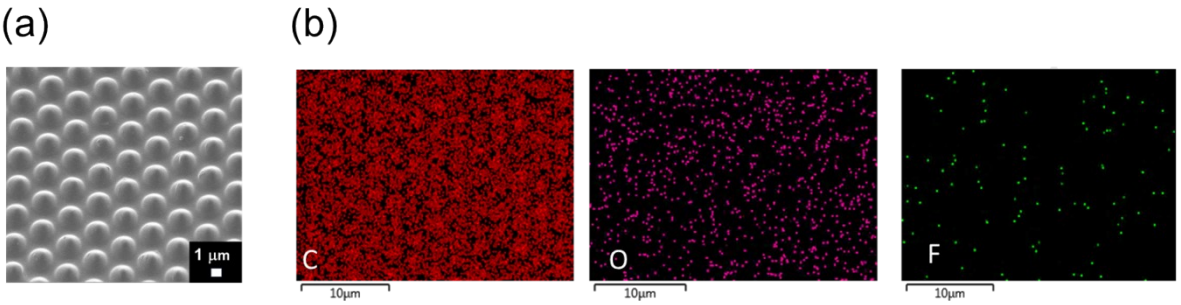


Fig. S3. (a) Scanning electron microscopy (SEM) images, (b) Energy dispersive X-ray spectroscopy (EDX) of the PTFE film.

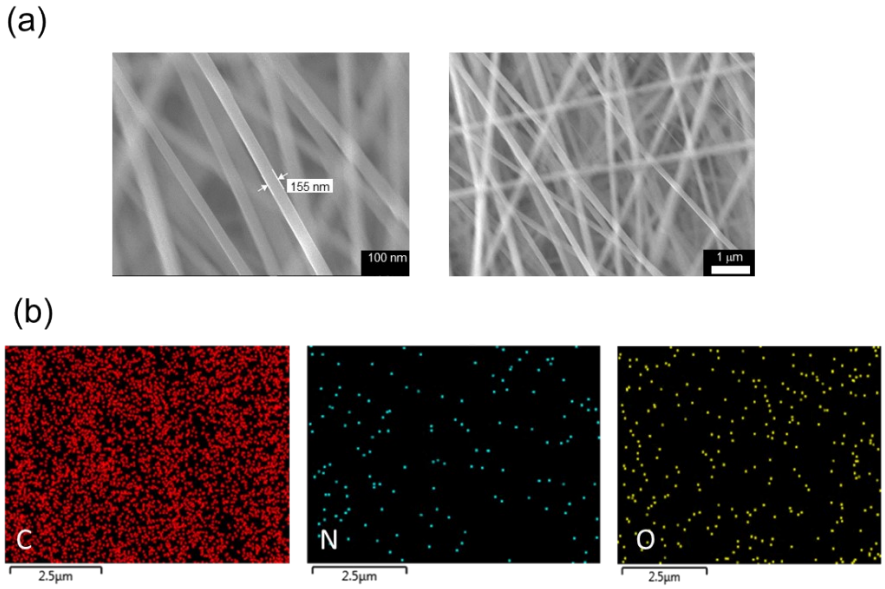


Fig S4. (a) Scanning electron microscopy (SEM) images, (b) Energy dispersive X-ray spectroscopy (EDX) of the nylon 6/6 nanofiber film.

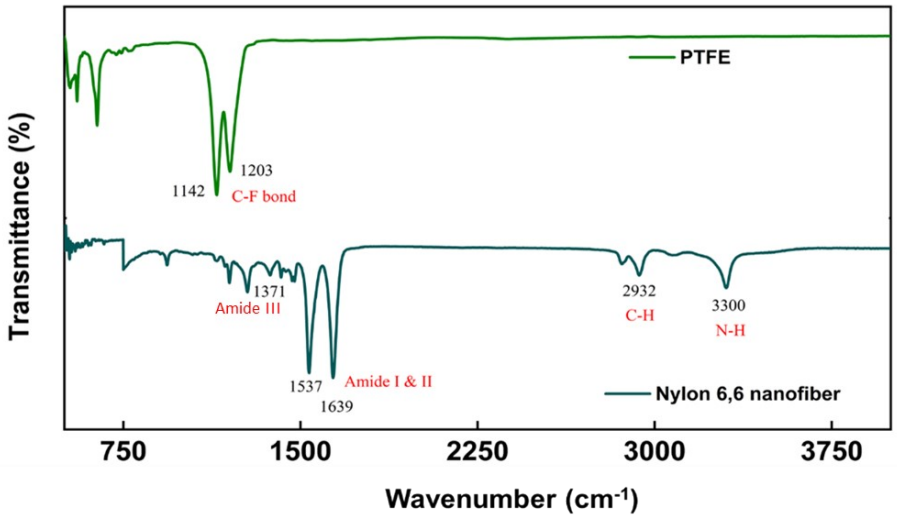


Fig. S5. Fourier transform infrared (FTIR) spectra of the PTFE and nylon 6/6 film.

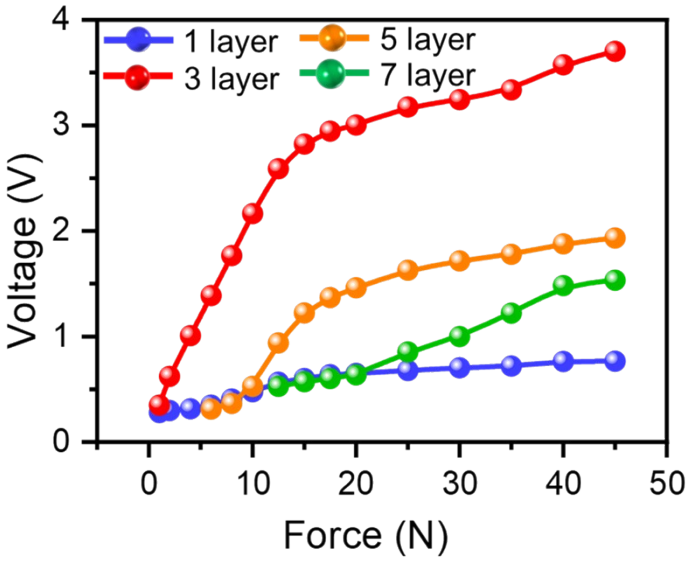


Fig. S6. Effect of the spacer layers on TENG output voltage (with force variations).

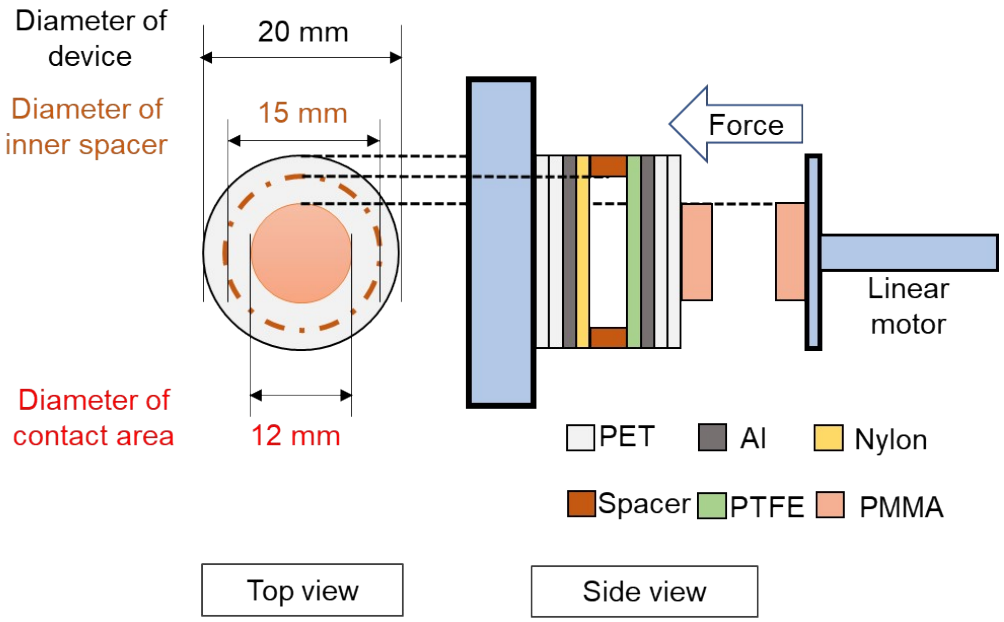


Fig. S7. The linear motor setup with the TENG sensor (top view and side view).

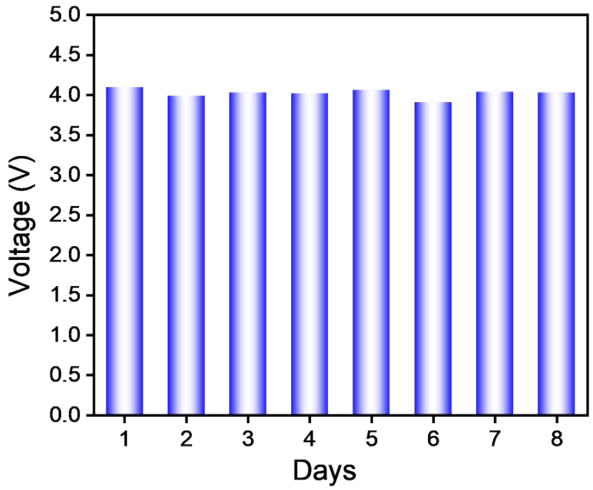


Fig. S8. Stability of the proposed TENG sensor across multiple days.

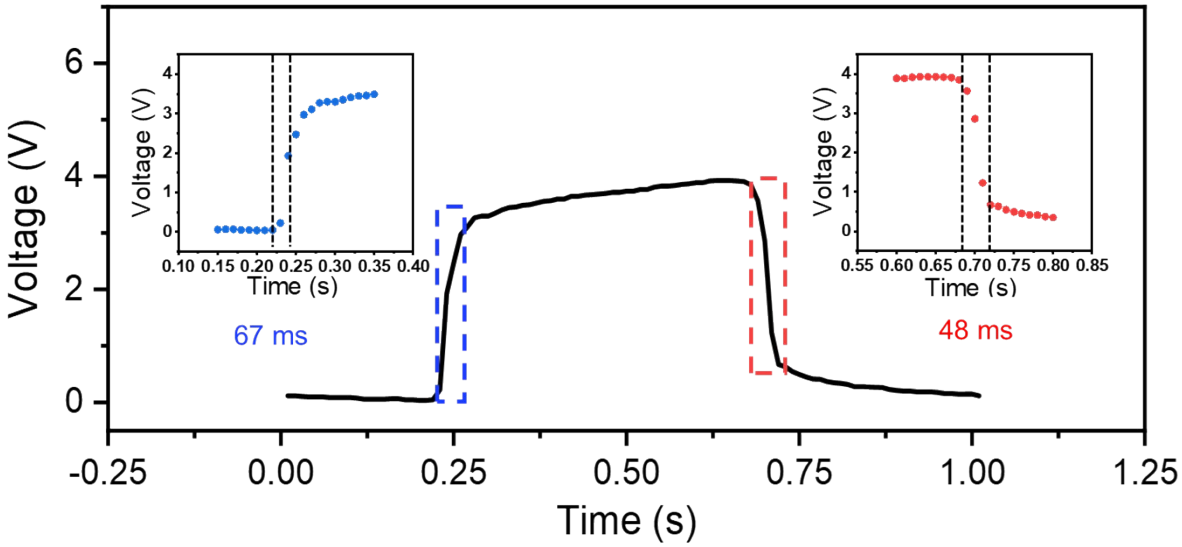


Fig. S9. Response and recovery times of the TENG sensor.

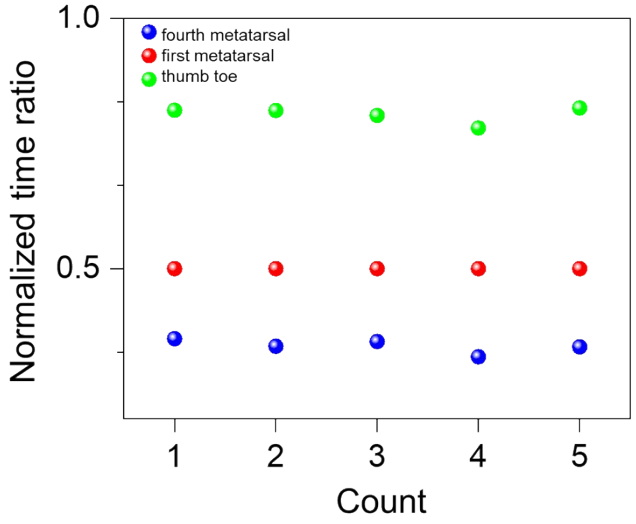
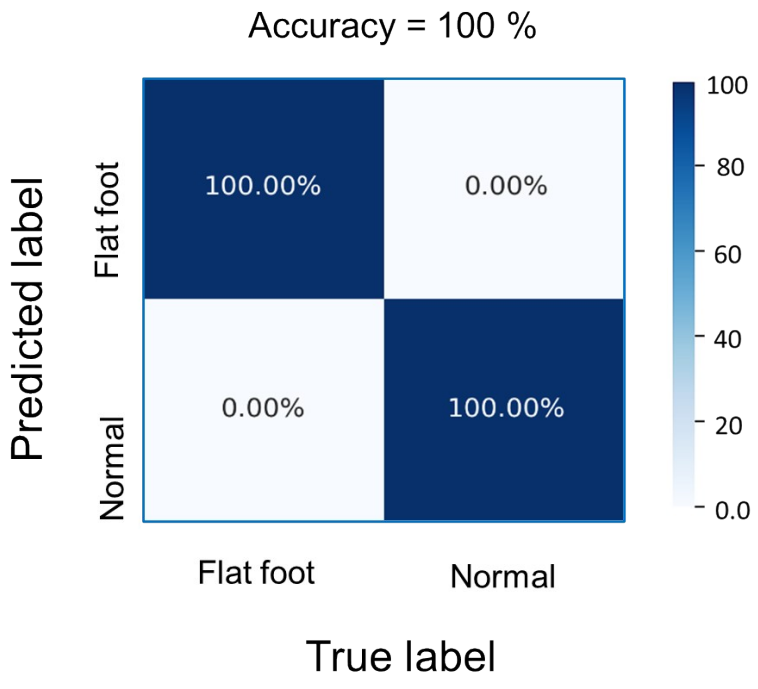


Fig. S10. Normalized time ratio of a healthy human gait.



\

Fig. S11. Confusion matrix for the normal arches from pes planus classification.

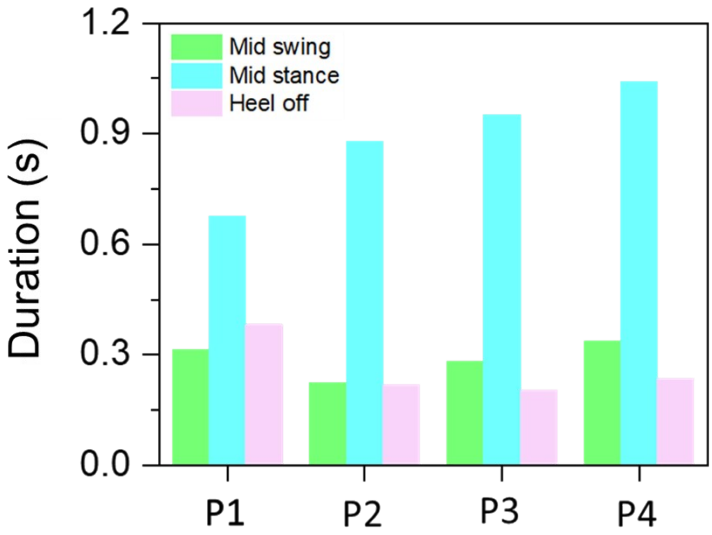


Fig. S12. Extracted average durations (left heel data) for gait phase events for 4 persons (P1, P2, P3, and P4).

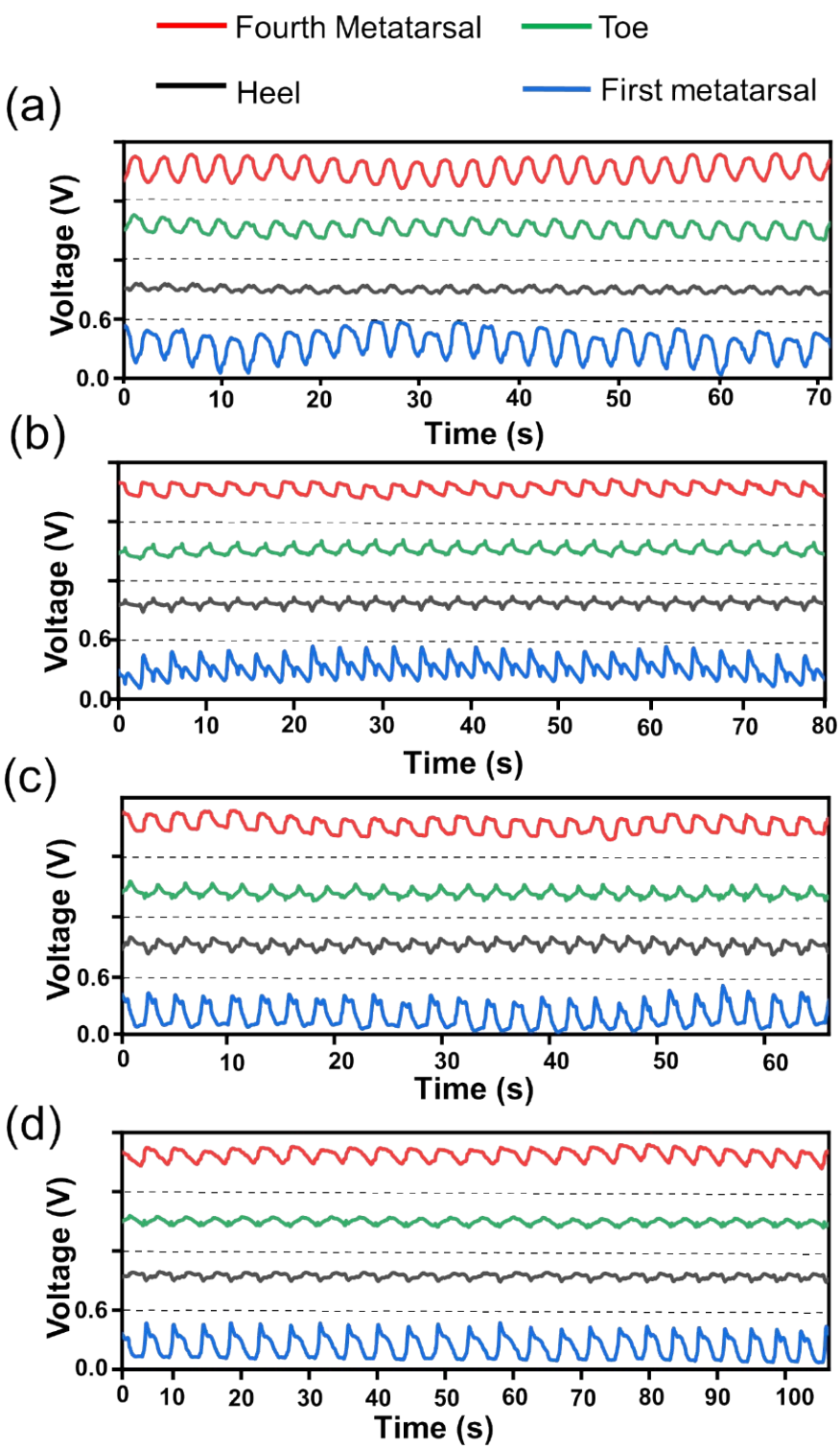


Fig. S13. Left foot raw data for 25 gait cycles for (a) person P1, (b) person P2, (c) person P3, and (d) person P4.

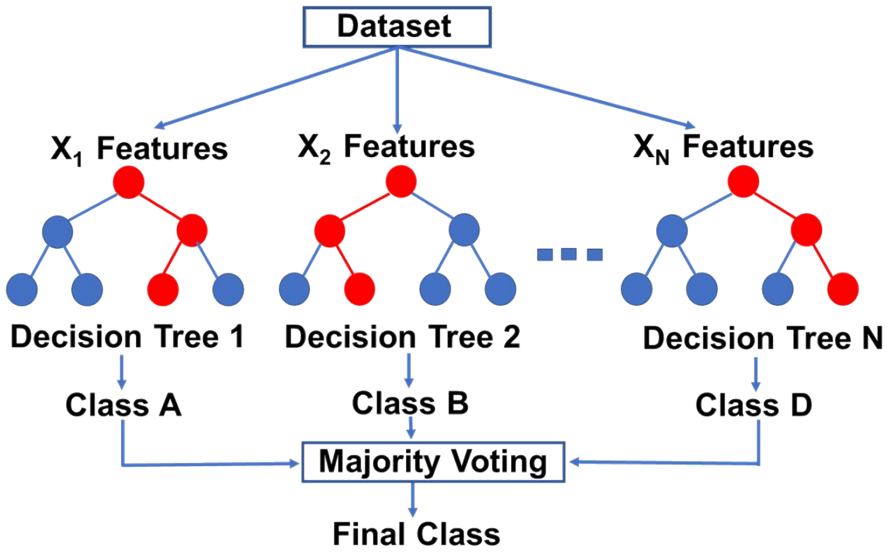


Fig. S14. Illustration of the random forest classifier machine learning model.

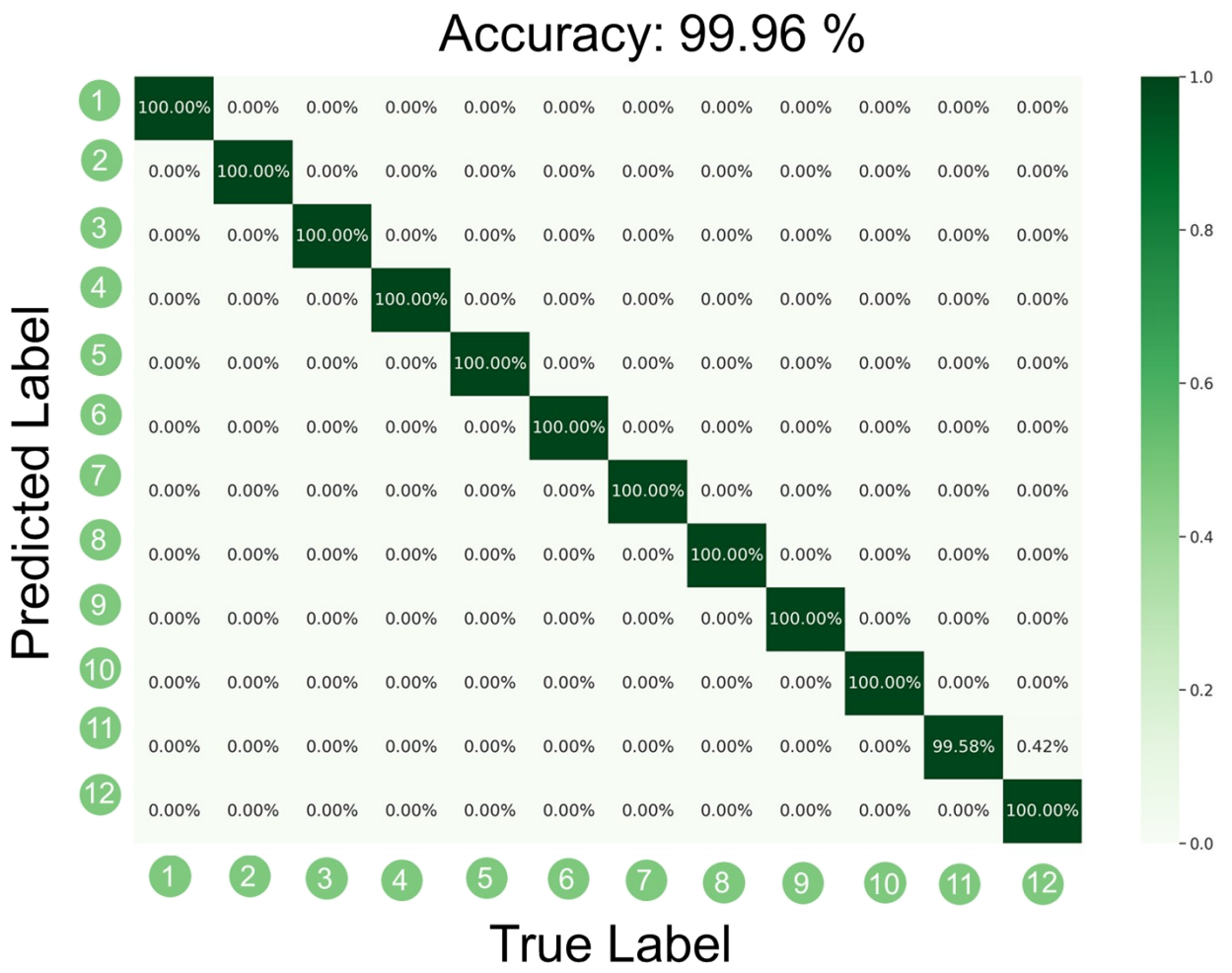


Fig. S15. Confusion matrix for the 12 rehabilitation exercises (half squat R1, standing on tiptoes R2, standing on heels R3, lifting legs up from the front side (while standing) R4, lifting legs up from the side direction (while standing) R5, lifting legs up from the rear side (while standing) R6, stepping leg up when sitting (on chair) R7, chair-assisted tiptoe squat down and standing up R8, chair sitting and standing with arms akimbo R9, standing lunges R10, stepping from the side directions R11, and walking on heels R12).

Fig. S16. Actual images of the 12 rehabilitation exercises.

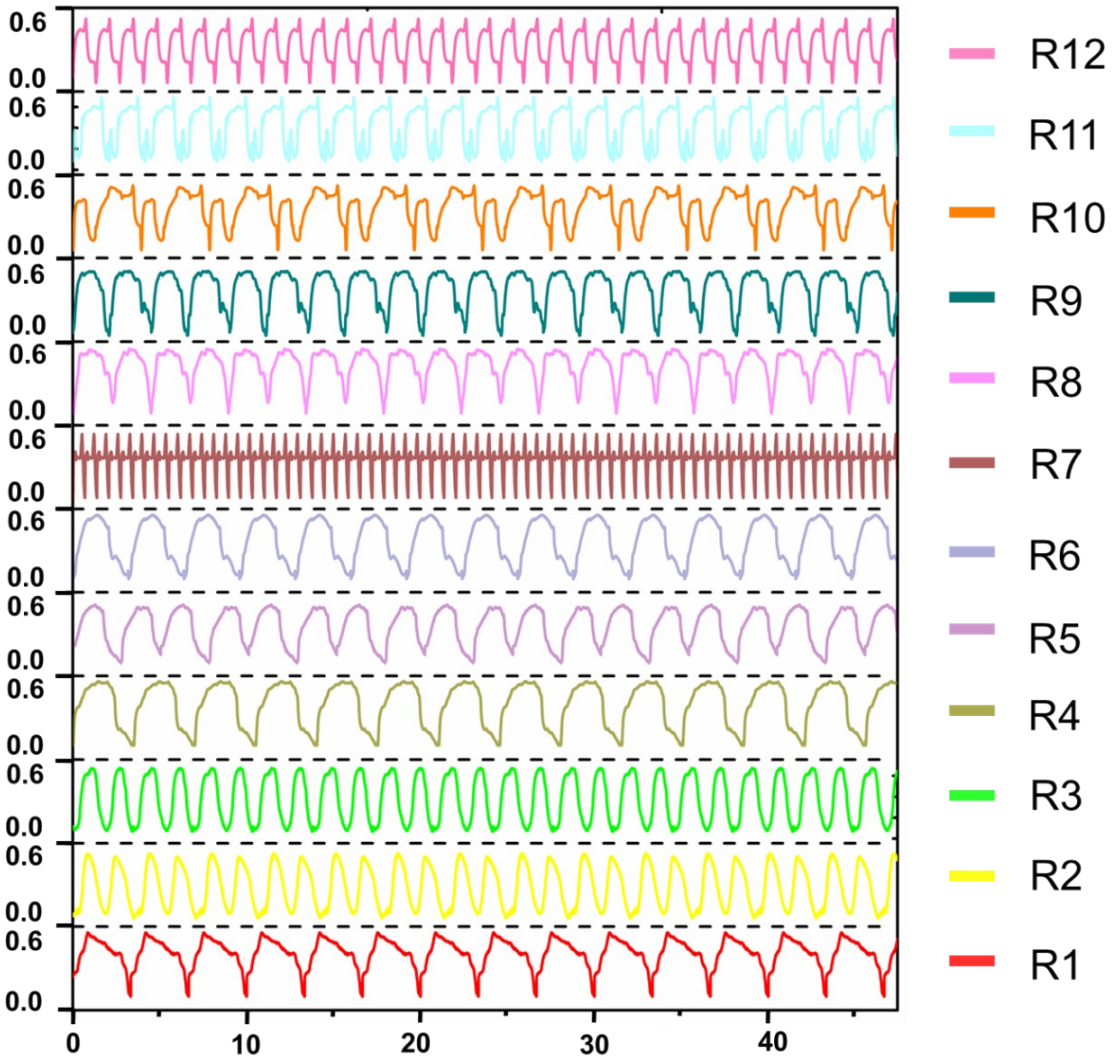


Fig. S17. The left heel data for 12 rehabilitation exercises (half squat R1, standing on tiptoes R2, standing on heels R3, lifting legs up from the front side (while standing) R4, lifting legs up from the side direction (while standing) R5, lifting legs up from the rear side (while standing) R6, stepping leg up when sitting (on chair) R7, chair-assisted tiptoe squat down and standing up R8, chair sitting and standing with arms akimbo R9, standing lunges R10, stepping from the side directions R11, and walking on heels R12).

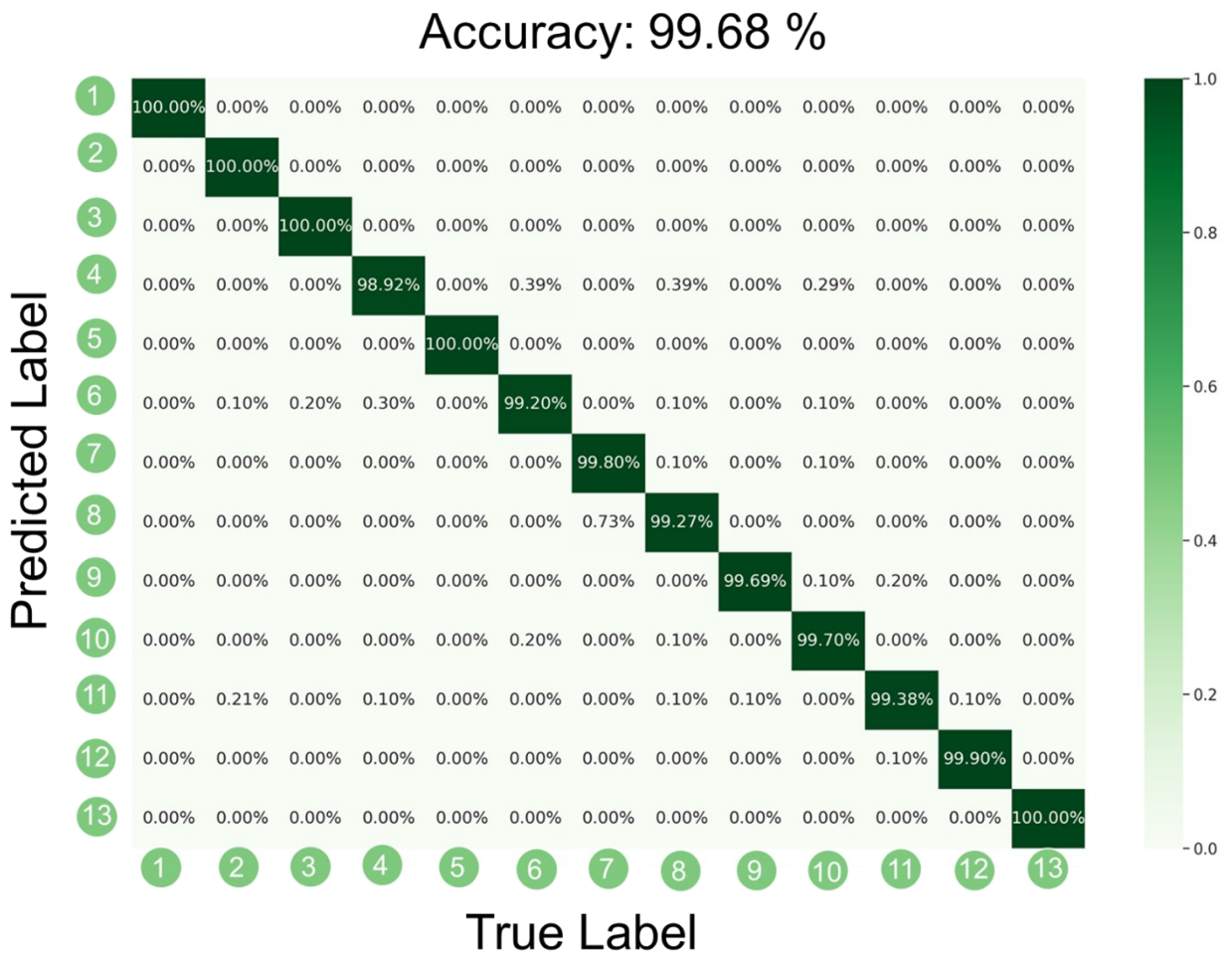


Fig. S18. The confusion matrix for 13 workout exercises (deep squat E1, mountain climbing E2, high knee E3, squat jacks E4, lateral lunges E5, rope jumping E6, walking up the stairs E7, walking down the stairs E8, burpee E9, jumping jacks E10, cross jumping E11, kicking hips E12, walking lunges E13).

Fig. S19. Actual images of the 13 workout exercises.

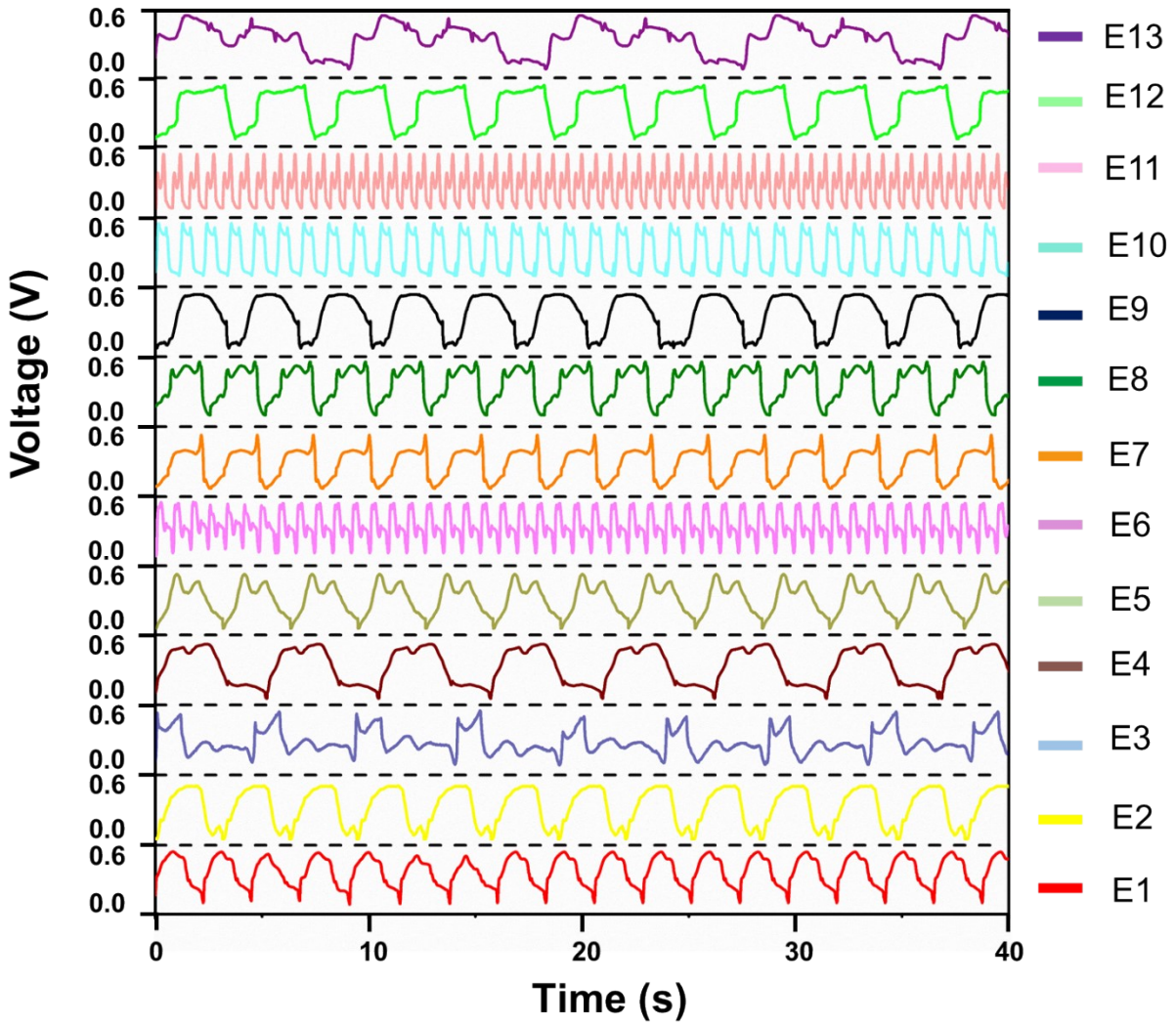


Fig. S20. The left heel data for 13 workout exercises (deep squat E1, mountain climbing E2, high knee E3, squat jacks E4, lateral lunges E5, rope jumping E6, walking up the stairs E7, walking down the stairs E8, burpee E9, jumping jacks E10, cross jumping E11, kicking hips E12, walking lunges E13)

Supporting Note S1 : Random forest classifier

Random forest is an advanced ensemble learning method designed to enhance the performance of predictive models by aggregating the outputs of multiple models. This technique is applicable to both classification and regression problems within machine learning. The foundation of a random forest classifier lies in its ensemble of decision trees, which serve as the fundamental units or base learners. Decision trees operate as binary structures, where data is repeatedly partitioned based on specific feature values, ultimately culminating in leaf nodes that signify the model's final predictions.[1,2] During the model training phase, each decision tree is constructed using a bootstrap sample—a randomly selected subset of data drawn with replacement from the original dataset as depicted in Fig. S14. This sampling strategy, combined with the presence of numerous trees, significantly reduces the risk of overfitting. Additionally, to introduce regularization, only a randomly chosen subset of features is evaluated at each decision point within the trees. This mechanism prevents any single tree from exerting undue influence on the ensemble's overall performance. For classification tasks, the random forest determines the final output by aggregating the majority of votes from individual trees. The algorithm's simplicity, adaptability, and effectiveness make it a popular choice. It is recognized for delivering high accuracy and being less susceptible to overfitting compared to standalone decision trees.

User Identification:

The voltage signals generated by four users (U1, U2, U3, and U4) are collected at a sampling frequency of 250 kHz. For user identification, the dataset comprises 60,000 samples (15,000 per class), with each sample containing data from eight sensors (four sensors per foot). The dataset is randomly split into an 80:20 ratio for training and testing. A random forest ensemble model with 15 estimators is trained on the training set, while the test set is utilized to evaluate model performance.

Flat foot identification: A total of 20,000 samples (10,000 per class), with each sample containing data from eight sensors, are taken into consideration. The dataset is randomly split into a 90:10 ratio for training and testing. A random forest ensemble model with 20 estimators is used for classification.

Rehabilitation and athletic workout classification:

The datasets for 12 rehabilitation exercises and 13 athletic workout exercises consist of 60,000 and 65,000 samples, respectively, with 5,000 samples per class. Each sample includes data from eight sensors evenly distributed between the left and right foot (four sensors per foot). Both datasets are randomly split into a 90:10 ratio, with 90% used for training and 10% for testing. A random forest ensemble model with 20 estimators is trained on the respective training sets, while the test sets are used to evaluate the classification performance.

References:

1. N. Altman, M. Krzywinski, *Nat. Methods*., 2017, **14**, 933–934.

2. L. Breiman, *Machine Learning*. , 2001, **45**, 5–32.